Interaction between Normative Systems and Cognitive Agents in Temporal Modal Defeasible Logic

Regis Riveret[1], Antonino Rotolo[1], Guido Governatori[3]

[1] CIRSIFD and Law Faculty
Via Galliera 3, 40121, Bologna, Italy
rriveret@cirsfid.unibo.it
rotolo@cirsfid.unibo.it
[2] University of Queensland
Brisbane, Queensland, QLD 4072, Australia
guido@itee.uq.edu.au

Abstract. While some recent frameworks on cognitive agents addressed the combination of mental attitudes with deontic concepts, they commonly ignore the representation of time. An exception is [1] that manages also some temporal aspects both with respect to cognition and normative provisions. We propose in this paper an extension of the logic presented in [1] with temporal intervals.

Keywords. Time, Norm, Temporal Modal Defeasible Logic

1 Introduction

A common approach in the agent literature for programming cognitive agents in a BDI (belief, desire, intention) framework is the use of rules to represent or manipulate agents mental attitudes. In addition to the three mental attitudes of beliefs, desires and intentions, some works include deontic concepts which are used to denote norms, commitments of social agents and social rationality [2][3][4][5][6]. However, these frameworks commonly ignore the representation of time. An exception is [1] that adopts the rule-based approach of [7][8][9] and extends it to accommodate temporal aspects. Time is integrated by pairing propositions with instants representing the time when propositions hold and by descriminating transient and persistent conclusions. Persistent conclusions persists through time until some interrupting events occurs. Such account of persistenccy is incomplete since some properties may ends without any external event. We propose in this paper an extension of the logic presented in [1] with temporal intervals so that persistent conclusions can ends by means of the occurence of an interrupting event or by means of a certain temporal reference.

The presented framework is based on Temporal Defeasible Logic (TDL) which is an umbrella expression to designate extensions of DL to capture time. TDL has been proved usefull in modelling temporal aspects of normative reasonning, such as temporalised normative provisions [10]; in addition, the notion of temporal viewpoint - the temporal position from which things are viewed- allows for a logical account of retroactibe norms and norm modidications [11].

Dagstuhl Seminar Proceedings 07122
http://drops.dagstuhl.de/opus/volltexte/2007/923

The layout of the paper is as follows. In section 2, we intoduce the general conceptual model behind the framework. Section 3 provides an outline of basic Defeasible Logic. Section 4 describes a variant of modal TDL that formalise the model of cognition.

2 Time, norms and mental attitudes

Our model aims to give account of some temporal aspects in respect to the combination of mental attitudes with deontic provisions. The starting point is the acknowledgement that on the first hand recent works shows that reasoning about agents can be embedded in frameworks based on non-monotonic logic, as the most interesting problems concerns the cases where the agent's mental attitudes are in conflict or when they are incompatible with some deontic provivions. On the other hand, in temporal setting, non-monotonicity can also be used to conclude that mental attitudes or deontic provivions persists up to some future time unless there is a reason for it not to perist. One can thus argue that a type of non-monotonicity concerns situation where mental attitudes are in conflict or when they are incompatible with some deontic provivions, while another type of non-monotonocity issues temporal aspects. Our model is based on these two types of non-monotonicity.

Accordingly, we adopt in this paper a system that follows the works of [7][8][9] which are themselves inspired by Bratman's analysis of so-called policy-based attitudes. In Bratman's view intentions are used to choose partial plans for realisation of a goal and have a close relation to mean-ends, whereas [7][8][9] intentions is related not only to means-ends but also to their consequences. This notion is partivularly relevant with deontic and normative notions, for example if we want to say that an agent is legally for A if the A is a side effect and if the agent did A with the intention to do A. [7][8][9] extends this policy-based approach to other attitudes and motivational factors as beliefs, intentions and oligations. An agent types correspond to the different ways thorugh which conflicts are detected and solved: a realistic agent thus corresponds to a conflict-resolution type in which beliefs override all other factors, while other agent types, such as simple minded, selfish or social ones adopt differet orders of overruling. Some substantial peculiarities makes it different from other frameworks such as BOID's. In particular,

1. the system develops a constructive account of these modalities; rules are thus meant to devise suitable logical conditions for introducing modalities; if so, rules may also contain modalised literals;
2. possible conversions of a modality into another can be accepted, as when the aplicability of rule leading to derive, for example, $\mathrm{OBL}p$ (p is obligatory) may permit, under appropriate conditions, to obtain $\mathrm{INT}p$ (p is intented).

[1] is on the same line of research of [7][8][9] and focus on some temporal aspects. [1] is based on Bratman's [12] which in his pursuit for a temporally extended rational agency exposed a principle that can be roughly stated as follows:

– At $t0$, agent A deliberates about what policy to adopt concerning a certain range of activities. On the basis of this deliberation agent A forms a general intention to φ in circumstances of type ψ.
– From $t0$ to $t1$, A retains this general intention.
– At $t1$, A notes that he/she is or will be in circumstance ψ at $t2$, where $t2 \geq t1$.
– Based on the previous steps, A forms the intention at $t1$ to φ at $t2$.

Given the temporal nature of Bratmans historical principle, and the idea that some intentions can be retained from one moment to another, [1] accounts for two types of temporal deliberations: transient deliberations, which hold only for an instant of time, and persistent deliberations, in which an agent is going to retain them unless some intervening event that forces the agent to reconsider her deliberation occurs. This event can be just a brute fact or it can be a modification of the policy of the agent. That is, persistent conclusions persists through time until some interrupting events occurs. Such account of persistenccy is incomplete since some properties may ends without any external event. We propose in this paper an extension of this model with temporal intervals so that persistent conclusions can ends by means of the occurence of an interrupting event or by means of a certain temporal reference. For example, intervals allows us to represent the span of time during which an obligation holds or a norm is in force.

Ordinarily, intervals are defined as sets of instants between two indicated instants. Here we deviate to this definition because of the non-homogeneity or transient character of events: if an event occurs in an interval conceived as a set of instants, then it would also occur in the set of instants that defines it and this would conflict with the transient characterisation of events. Hence, we define an interval as a pair of instants of the form $[t_i, t_f]$. Intervals are usually denoted by T (plus eventual subscript). We identify two subtsets of interval to differentiate intervals in which an associated property holds at any instant between the boundaries and intervals in which an associated property holds at least one instant between the boundaries. We shall call the firsts A-interval and the seconds B-intervals. A-intervals are represented by expressions of the form $\overline{[t_i, t_f]}$ and are usually denoted by $\overline{T}$ while B-intervals are represented by expressions of the form $\widehat{[t_f, t_f]}$ and denoted by $\widehat{T}$. If the wide hat or the line over an interval is omitted then it is either an A-interval or a B-interval.

Mental attitudes and normative provisions are related to temporal references and the passage of time allows change of these elements. This is in accordance with the commonly accepted opinion that in a static system where nothing changes, the temporal dimension does not provide more understanding.
Our references (intervals) allows us to temporalise literals and rules. In its simplest form, a temporal literal is an expression of the form $l{:}T$ where l is a literal and T is either an A-interval or a B-interval. Intuitively, $l{:}\overline{T}$ means that l holds for all instants between the boundaries of $\overline{T}$ while $l{:}\widehat{T}$ means that l holds for at least an instant between the boundaries of $\widehat{T}$. For example, $major(bob){:}\overline{[1973, max]}$ means that Bob is major in 1973 and later. Similarly, rules are temporalised by associating to it a time interval,and so a temporal rule is an expression of the form:

$(r\text{: } a_1{:}T_1 \ldots a_n{:}T_n \hookrightarrow b{:}\overline{T}){:}\overline{T_r}$

The time labels allow us to deal formally with the different temporal dimensions of a normative system. The temporal intervals labelling the antecedent of a rule, the consequent of the rule and the overall rule are interpreted respectively as the intervals of efficacy, applicability and time of force of the represented provision. These diferrent temporal dimensions are in line with the legal temporal model developed in [13]. That allows us to give an accurate account of temporal aspects of norms and therefore to be consistent with legal principles. Note that the interval $\overline{T_r}$ labelling the entire rule is an A-interval because the force of a provision is generally an homogeneous property. Similarly, we constraint the interval labelling the literal in the head of the rule to be an A-interval. Intervals in the body can be A-intervals or B-intervals. An example of a temporal rule is:

$$(r\colon\ born(X){:}[t,t] \rightarrow major(X){:}\overline{[t+18, max]}){:}\overline{[1970, max]}$$

This rule formalises the provision in force in 1970 and later that somebody get its majority at 18 years old.

Consequently of the different temporal dimensions, a conclusion can be associated to two temporal intervals. The first interval is the interval with which the consequent of the rule is labelled while the second interval corresponds to the time of force interval associated to the rule. We represent such temporalisation of conclusion by concatenation of intervals by means of the symbol ':' and we call such concatenation chain of viewpoints. For example. giving the rule r and the fact that Bob was born in 1960, then one can conclude $major(bob){:}[1978, max]{:}[1970, max]$, that is, Bob is major in 1978 (and later) from somebody reasoning in 1970 (and later). Chain of viewpoints are of the upmost importance when one has to deal with the retroactivity of norms. Retroactivity occurs when the effects of a rule r apply to an interval $\overline{[t_i, t_f]}$ which begins before the interval $[t'_i, t'_f]$ attached to the antecedent of r, that is, $t_i < t'_i$. Another case of retroactivity is when the consequence of a rule r' applies at an interval $\overline{[t_i, t_f]}$ but r' is in force in $\overline{[t_{ri}, t_{rf}]}$ such that $t_i < t_{ri}$. For an illustration of the utility of chain of viewpoints with respects to retroactivity, consider the following rules:

$$(r1\colon\ Income > 90{:}[1\widehat{Mar06, max}] \Rightarrow_{\mathrm{OBL}} \neg Tax{:}\overline{[1Jan06, max]}){:}\overline{[15Jan06, max]}$$

$$(r2\colon\ Income > 100{:}[1\widehat{Mar06, max}] \Rightarrow_{\mathrm{OBL}} Tax{:}\overline{[1Jan06, max]}){:}\overline{[1Apr06, max]}$$

Rule $r1$ states that if the income of a person is in excess of ninety thousand as of 1st March 2006 then she has not to pay the tax from 1st January 2006 with the policy being in force from 15 January 2006. This means that the norm becomes part of the tax regulation from 15 January 2006. Accordingly, the policy covers tax returns lodged after 15 January 2006. The second rule, in force from 1st April 2006, establishes a tax returns lodged after 1st April 2006. These two rules illustrate the concept of viewpoints. Consider that the conditions in the antecedent of both rules hold, then one would derive $\neg Tax{:}\overline{[1Jan06, max]}{:}\overline{[15Jan06, max]}$ but $Tax{:}\overline{[1Jan06, max]}{:}\overline{[1Apr06, max]}$, that is, from if one reason from a point of view between the 15 January and the 1st April then the tax is due while if one reason from a point of view later than the 1st April then no tax is due.

Even though trivial cases of the phenomenon of retroactivity are captured by rules such as $r1$ and $r2$, we should be able to detect retroactivity also in other scenarios, where normative effects are in fact applied retroactively to some conditions as a result of complex arguments that involve many rules. This problem is of great importance not only because the designer of a normative system may have the goal to state retroactive effects in more articulated scenarios, but also because she should be able to check whether such effects are not obtained when certain regulations regard matters for which retroactivity is not in general permitted. This is the case of criminal law, where the principle -Nullum crimen, nulla poena sine praevia lege poenali- is valid.

3 Defeasible Logic

The legal temporal model points out the importance of uncertainty due to the addition of new premises that can invalidate formerly derivable consequences. This means that temporal cognition must proceed on the basis of non-monotonic reasoning. Non-monotonic reasoning is supported by a number of non-monotonic logics. Among these, DL [14][15][16] is based on a logic programming-like language and it is a simple, efficient but flexible non-monotonic formalism capable of dealing with many different intuitions of non-monotonic reasoning. An argumentation semantics exists [17] that makes its use possible in argumentation systems. DL has a linear complexity [18] and also has several efficient implementations [19].

A Defeasible Logic theory is a structure $D = (F, R, \prec)$ where F is a finite set of facts, R a finite set of rules, and $\prec$ a superiority relation on R. Facts are indisputable statements, for example, "Bob is a minor," formally written as $minor(bob)$. Rules can be strict, defeasible, or defeaters. Strict rules are rules in the classical sense; whenever the premises are indisputable, so is the conclusion. An example of a strict rule is "Minors are persons," formally written as $r1$: $minor(X) \rightarrow person(X)$. Defeasible rules are rules that can be defeated by contrary evidence. An example of a defeasible rule is "Persons have legal capacity"; formally, $r2$: $person(X) \Rightarrow haslegalcapacity(X)$. Defeaters are rules that cannot be used to draw any conclusion. Their only use is to prevent some conclusions by defeating some defeasible rules. An example of this kind of rule is "Minors might not have legal capacity," formally expressed as $r3$: $minor(X) \leadsto \neg haslegalcapacity(X)$. The idea here is that even if we know that someone is a minor, this is not sufficient evidence for the conclusion that he or she does not have legal capacity. The superiority relation between rules indicates the relative strength of each rule. That is, stronger rules override the conclusions of weaker rules. For example, if $r3 \succ r2$, then the rule $r3$ overrides $r2$, and we can derive neither the conclusion that Bob has legal capacity nor the conclusion that he does have legal capacity.

Given a set R of rules, we denote the set of all strict rules in R by R_s, the set of defeasible rules in R by R_d, the set of strict and defeasible rules in R by R_{sd}, and the set of defeaters in R by R_{dft}. $R[q]$ denotes the set of rules in R with consequent q. In the

following $\sim p$ denotes the complement of p, that is, $\sim p$ is $\neg p$ if p is an atom, and $\sim p$ is q if p is $\neg q$. For a rule r we will use $A(r)$ to indicate the body or antecedent of the rule and $C(r)$ for the head or consequent of the rule. A rule r consists of its antecedent $A(r)$ (written on the left; $A(r)$ may be omitted if it is the empty set), which is a finite set of literals; an arrow; and its consequent $C(r)$, which is a literal. In writing rules we omit set notation for antecedents.
Conclusions are tagged according to whether they have been derived using defeasible rules or strict rules only. So, a conclusion of a theory D is a tagged literal having one of the following four forms:

$+\Delta q$ meaning that q is definitely provable in D.
$-\Delta q$ meaning that q is not definitely provable in D.
$+\partial q$ meaning that q is defeasibly provable in D.
$-\partial q$ meaning that q is not defeasibly provable in D.

These different notions of provability come of use here because they enable the system to label a suggestion as stronger or weaker depending on the kind of proof associated with it. Provability is based on the concept of a derivation (or proof) in D. A derivation is a finite sequence $P = (P(1),...,P(n))$ of tagged literals. Each tagged literal satisfies some proof conditions. A proof condition corresponds to the inference rules that refer to one of the four kinds of conclusions we have mentioned above. $P(1..n)$ denotes the initial part of the sequence P of length n. We state below the conditions for defeasibly derivable conclusions:

If $P(i+1) = +\partial q$ then
(1) $+\Delta q \in P(1..i)$ or
(2) (2.1) $\exists r \in R_{sd}[q] \forall a \in A(r) : +\partial a \in P(1..i)$ and
(2.2) $-\Delta \sim q \in P(1..i)$ and
(2.3) $\forall s \in R[\sim q]$ either
(2.3.1) $\exists a \in A(s) : -\partial a \in P(1..i)$ or
(2.3.2) $\exists t \in R_{sd}[q]$ such that
$\forall a \in A(t) : +\partial a \in P(1..i)$ and $t \succ s$.

If $P(i+1) = -\partial q$ then
(1) $-\Delta q \in P(1..i)$ and
(2) (2.1) $\forall r \in R_{sd}[q]\ \exists a \in A(r) : -\partial a \in P(1..i)$ or
(2.2) $+\Delta \sim q \in P(1..i)$ or
(2.3) $\exists s \in R[\sim q]$ such that
(2.3.1) $\forall a \in A(s) : +\partial a \in P(1..i)$ and
(2.3.2) $\forall t \in R_{sd}[q]$ either
$\exists a \in A(t) : -\partial a \in P(1..i)$ or $t \not\succ s$.

Informally, a defeasible derivation for a provable literal consists of three phases: First, we propose an argument in favour of the literal we want to prove. In the simplest case, this consists of an applicable rule for the conclusion (a rule is applicable if its antecedent has already been proved). Second, we examine all counter-arguments (rules for the opposite conclusion). Third, we rebut all the counter-arguments (the counter-argument is weaker than the pro-argument) or we undercut the (some of the premises of the counterargument are not provable).

4 Temporal Modal Defeasible Logic

Defeasible Logic allows us to deal with incomplete information but as such does not provide any mean to deal with modalities and temporal aspects. Temporal Modal Defeasible Logic is an umbrella expression to designate possible extensions of Defeasible Logic to capture modalities and time. We present in this section an extension of [1] with intervals as exposed in the temporal model (see Section 2).

4.1 Modal Domain

The combination of mental attitudes and obligations are framed in Defeasible Logic following the works of [7][8][9] and capture some basic facets of the modal notions of knowledge, intentions, action and obligation.

To extend DL woth modal opators, new types of rules relative to modal operator are introduced: arrows of the rules are labelled by the different modalities we want to deal with. This solution leads to distinguishing different modes through which the literals can be derived using rules. How such types of derivation are related to the introduction of the corresponding modalised literals can be expressed as follows: if $X \in \{\text{KNOW}, \text{INT}, \text{ACT}, \text{OBL}\}$, then

$$\frac{\Gamma \quad \Gamma \Rightarrow_X \psi}{\Gamma \mid\!\sim X\psi} \text{ MI}$$

As we will see, we do make an exception when rules for knowledge are concerned since we will state that $X \in \{\text{INT}, \text{OBL}, \text{ACT}\}$. The reason for this is that we assume that beliefs are conceived of as the knowledge the agent has of the environment, and so they are used by the agent to make inferences about how the world is: in this perspective, belief conclusions correspond to factual knowledge and do not need to be modalised. But besides this exception, which can be removed if required, schema MI captures the basic logical behaviour of our modal rules. Notice, also, that actions are successful and intentional and so, when $\text{ACT}\psi$ is derived, this also implies that ψ and $\text{INT}\psi$ are the case.

Other relations between modalities are captured by means of *rule conversions* and *conflicts*.

The notion of *rule conversion* permits to use rules for a modality X as they were for another modality Y. Suppose that a rule of a specific type is given and also suppose that all the literals in the antecedent of a rule are provable in one and the same modality, then it is arguable that the conclusion of the rule inherits the modality of the antecedent. An example can help us illustrate the notion of conversion. Consider the following formalisation of the Yale Shooting Problem.
load:$\overline{[t,t]}$, shoot:$\overline{[t,t]}$ $\Rightarrow_{\text{KNOW}}$ kill:$\overline{[t,t]}$
This rule encodes the knowledge of an agent that knows that loading the gun with live

ammunitions, and then shooting will kill her friend. This example clearly shows that the qualification of the conclusions depends on the modalities relative to the individual acts "load" and "shoot". In particular, if we obtain that the agent intends to load and to shoot the gun (INT($load$), INT($shoot$)), then, since she knows that the consequence of these actions is the death of her friend, she intends to kill him. However, if shooting was not intended, then we have prima facie to say that killing, too, was not intentional.

To define the admitted conversions we introduce a binary relation Convert over the modalities of the language. When we write Convert$(\mathrm{KNOW},\mathrm{OBL})$ this means that a knowledge rule r can be used to derive an obligation (of course, provided that all its antecedents are derived as obligations): r can thus be converted into a rule for intention.

Conflicts play an important role in the current context and it is crucial to establish criteria for detecting and solving conflicts between the different components which characterise the cognitive profiles of agent's deliberation, and, above all between mental states and normative provisions. Conflicts are detected and solved by a similar strategy than basic DL, i.e, by following a pattern such that (i) in a first phase an argument supporting the conclusion is advanced (ii) in the second phase any possibble attack are considers, and (iii) finally the counter-attack for each attack.

Thus, for the purpose of this paper, we introduce a ternary relation Attack over the set of modalities that defines which types of rules are in conflict and which are the stronger ones. For example, if we write Attack$(\mathrm{OBL},\mathrm{INT},\mathrm{ACT})$ this means that, in the reasoning pattern illustrated above, obligations in general override intentions, which in turn override actions.

The relation Attack is explicitly linked to that of agent type. Classically, agent types are characterised by stating conflict resolution types in terms of orders of overruling between rules [3,7,9,8]. In this perspective, agent types are meaningful within a non-monotonic setting and are nothing but general strategies to detect and solve conflicts between the different components of the cognitive profiles of agent's deliberation. In [3] 24 possible types are identified while, in [8], based on a different framework, 20 combinations are proposed. Typically, rational agents are assumed to be at least *realistic*: a realistic agent, in fact, is such that rules for knowledge override all other components. If the realistic condition is abandoned, we may have various forms of wishful thinking. Given the minimal assumption that a rational agent should be realistic, we may further constrain agent's deliberation in order not to violate obligations: a *social agent* type requires that obligations are stronger than the other motivational components with the exception of beliefs. Other agent types can be specified, for which see [7][8][9].

4.2 Temporal domain

Approaches in temporal resonning are traditionally based on either instants (van Benthem, 1991), intervals (Allen, 1984) or both by representing one through the other (Allen and Hayes, 1989). We represent intervals by means of instants. Formally, we consider a totally ordered discrete set Temp of points of time termed -instants- and over it the order relation $> \subseteq Temp \times Temp$. We usually denotes the variables ranging over the members of Temp by t and its subscripts.

Ordinarily, intervals are defined as sets of instants between two indicated instants. Here we deviate to this definition because of the non-homogeneity or transient character of events: if an event occurs in an interval conceived as a set of instants, then it would also occur in the set of instants that defines it and this would conflict with the transient characterisation of events. Hence, we define an interval as a pair of instants. Formally, an interval is a member of the set $Int = \{[t_1,t_2] \in Temp \times Temp | t_1 \le t_2\}$. As can be noted, this defintion allows punctual intervals, i.e., intervals of the form $[t,t]$. Among the set Int, we identify two subtsets of interval to differentiate intervals in which an associated property holds at any instant between the boundaries and intervals in which an associated property holds at least one instant between the boundaries. We shall call the first A-interval and the seconds B-intervals. The set of A-intervals is denoted AInter while the set of B-intervals is denoted BInter. We have $\mathrm{AInter} \cap \mathrm{BInter} = \emptyset$ and $\mathrm{AInter} \cup \mathrm{BInter} = Int$. We shall usually denote intervals by T, A-intervals by $\overline{T}$ and B-intervals by $\widehat{T}$ (plus eventual subscripts).

As explained in section 2, a conclusion can be associated to two temporal intervals consequently of the different temporal dimensions. The first interval is the interval of applicability with which the consequent of the rule is labelled while the second interval corresponds to the time of force interval associated to the rule. Each interval can be assimilated to temporal russian-dolled viewpoints from which conclusions are considered. We represent such temporalisation of conclusion by concatenation of intervals by means of the symbol ':' and we call such concatenation chain of viewpoints. Chain of viewpoints are denoted by V (plus eventual subscripts).

Temporal calculi are driven by operators over intervals. In the literature, one can find many relations that hold between intervals. For example, (Allen 1984) proposes an algebra of intervals with thirteen mutually exclusive relations between two intervals. For our purpose, we limit the set of relations to -subinterval- denoted $\sqsubseteq$, over, -start in- denoted si, -start before end- denoted sbe and -start before start- denoted sbs.

Definition 1. *Let two intervals* $T = [t_i, t_f]$ *and* $T' = [t'_i, t'_f]$,

$T \sqsubseteq T'$ *iff* $t'_{mi} \le t_{mi}$ *and* $t_{mf} \le t'_{mf}$.

$si(T,T')$ *iff* $t'_i \le t_i \le t'_f$.

$sbe(T,T')$ *iff* $t_i \le t'_f$.

$sbs(T,T')$ *iff* $t_i \le t'_i$.

$over(T,T')$ *iff* $t'_i \le t_i \le t'_f$ *or* $t'_i \le t_f \le t'_f$ *or* $t_i \le t'_i \le t_f$.

Note that $T \sqsubseteq T'$, $\mathrm{si}(T,T')$ or $\mathrm{sbe}(T,T')$ implies $over(T,T')$, that $T \sqsubseteq T'$ implies $\mathrm{si}(T,T')$ and that $over(T,T')$ implies $over(T',T)$.
In order to lighten the paper, we may use the abbreviation consisting in placing chain of view points as arguments of the previous operators, such that for example,

- $T \sqsubseteq T' : T''$ stands for $T \sqsubseteq T'$ and $T \sqsubseteq T''$.
- $T : T' \sqsubseteq T'' : T'''$ stands for $T \sqsubseteq T''$ and $T' \sqsubseteq T'''$.
- $T : T' \sqsubseteq T''$ stands for $T \sqsubseteq T''$ and $T' \sqsubseteq T''$.

and similarly for other operators.

Finally, we consider the functions start() and end() that returns respectively the lower bound and upper bound of an interval.

4.3 The Language

A temporal defeasible agent theory consists of a discrete totally ordered set of instants of time Temp, a set of *facts* or indisputable statements, four sets of rules for knowledge, intentions, intentional actions, and obligations, and a *superiority relation* $>$ among rules saying when a single rule may override the conclusion of another rule. For $X \in \{\mathrm{KNOW}, \mathrm{INT}, \mathrm{ACT}, \mathrm{OBL}\}$, a *strict rule* is an expression of the form $\phi_1, \ldots, \phi_n \rightarrow_X \psi$ such that whenever the premises $\phi_1, \ldots, \phi_n$ are indisputable so is the conclusion ψ. A *defeasible rule* is an expression of the form $\phi_1, \ldots, \phi_n \Rightarrow_X \psi$ whose conclusion can be defeated by contrary evidence. An expression $\phi_1, \ldots, \phi_n \leadsto_X \psi$ is a *defeater* used to defeat some defeasible rules by producing evidence to the contrary. It is worth noting that modalised literals can occur only in the antecedent of rules: the reason of this is that the rules are used to derive modalised conclusions while we do not conceptually need to iterate modalities. This limitation makes the system more manageable.

Definition 2 (Language). *Let* Temp *a discrete totally ordered set of instants of time,* Prop *be a set of propositional atoms,* $\mathrm{Mod} = \{\mathrm{KNOW}, \mathrm{INT}, \mathrm{ACT}, \mathrm{OBL}\}$ *be the set of modal operators, and* Lab *be a set of labels. The sets below are the smallest sets closed under the following rules:*

Literals

$$\mathrm{Lit} = \mathrm{Prop} \cup \{\neg p | p \in \mathrm{Prop}\}$$

If q is a literal, $\sim q$ denotes the complementary literal (if q is a positive literal p then $\sim q$ is $\neg p$; and if q is $\neg p$, then $\sim q$ is p);

Modal Literals

$$\mathrm{ModLit} = \{Xl, \neg Xl | l \in \mathrm{Lit}, X \in \{\mathrm{INT}, \mathrm{ACT}, \mathrm{OBL}\}\};$$

Intervals

$$\mathrm{Inter} = \{T = [t1, t2] | t1, t2 \in \mathrm{Temp}, t1 \leq t2\};$$

A-Intervals

$$\mathrm{AInter} = \{\overline{T} = \overline{[t1, t2]} | t1, t2 \in \mathrm{Temp}, t1 \leq t2\};$$

B-Intervals

$$\mathrm{BInter} = \{\widehat{T} = \widehat{[t1, t2]} | t1, t2 \in \mathrm{Temp}, t1 \leq t2\};$$

Chain of Viewpoints

$$\mathrm{ChainView} = \{V = T1, V' = T1 : T2 | T1, T2 \in \mathrm{AInter} \cup \mathrm{BInter}\};$$

Temporal Literals

$$\mathrm{TempLit} = \{l : T | l \in \mathrm{Lit}, T \in \mathrm{AInter} \cup \mathrm{BInter}\};$$

Multi-Temporal Literals

$$\mathrm{MTempLit} = \{l : V | l \in \mathrm{Lit}, V \in \mathrm{ChainView}\};$$

Temporal Modal literals

$$\text{TempModLit} = \{Xl : T | Xl \in \text{ModLit}, T \in \text{AInter} \cup \text{BInter}\};$$

Multi-Temporal Modal literals

$$\text{MTempModLit} = \{Xl : V | Xl \in \text{ModLit}, V \in \text{ChainView}\};$$

Temporal Rules

$$\begin{aligned}
&\text{Rule}_s = \{(r : \phi_1, \ldots, \phi_n \rightarrow_X \psi) : T |\\
&\quad r \in \text{Lab}, A(r) \subseteq \text{TempLit} \cup \text{TempModLit}, X \in \text{Mod}, \psi \in \text{TempLit}, T \in \text{AInter}\}\\
&\text{Rule}_d = \{(r : \phi_1, \ldots, \phi_n \Rightarrow_X \psi) : T |\\
&\quad r \in \text{Lab}, A(r) \subseteq \text{TempLit} \cup \text{TempModLit}, X \in \text{Mod}, \psi \in \text{TempLit}, T \in \text{AInter}\}\\
&\text{Rule}_{dft} = \{(r : \phi_1, \ldots, \phi_n \leadsto_X \psi) : T |\\
&\quad r \in \text{Lab}, A(r) \subseteq \text{TempLit} \cup \text{TempModLit}, X \in \text{Mod}, \psi \in \text{TempLit}, T \in \text{AInter}\}\\
&\text{Rule} = \text{Rule}_s \cup \text{Rule}_d \cup \text{Rule}_{dft}
\end{aligned}$$

We use some abbreviations: $A(r)$ denotes the set $\{\phi_1, \ldots, \phi_n\}$ of *antecedents* of the rule r, and $C(r)$ to denote the *consequent* ψ of the rule r. We use also superscript for mental attitude, subscript for type of rule, and $\text{Rule}[\phi]$ for rules whose consequent is ϕ. If one does not refer to the content of the rule, a temporal rule can be written as $r{:}\overline{T}$ where r is the label of the rule and $\overline{T}$ is a temporal interval.

Definition 3 (Defeasible Agent Theory). *A defeasible agent theory is a structure*

$$D = (Temp, F, R^{\text{KNOW}}, R^{\text{INT}}, R^{\text{ACT}}, R^{\text{OBL}}, >, \mathscr{C}, \mathscr{V})$$

where

- Temp *a discrete totally ordered set of instants of time;*
- $F \subseteq \text{TempLit} \cup TempModLit$ *is a finite set of facts;*
- $R^{\text{KNOW}} \subseteq \text{Rule}^{\text{KNOW}}$, $R^{\text{INT}} \subseteq \text{Rule}^{\text{INT}}$, $R^{\text{ACT}} \subseteq \text{Rule}^{\text{ACT}}$, $R^{\text{OBL}} \subseteq \text{Rule}^{\text{OBL}}$ *are four finite sets of rules such that each rule has a unique label;*
- $> \subseteq R^{\text{KNOW} \cup \text{INT} \cup \text{ACT} \cup \text{OBL}} \times R^{\text{KNOW} \cup \text{INT} \cup \text{ACT} \cup \text{OBL}}$ *is an acyclic superiority relation.*
- $\mathscr{C} \subseteq \{Convert(X,Y) | X, Y \in \text{Mod}\}$ *is a set of conversions.*
- $\mathscr{V} \subseteq \{Attack(X,Y,Z) | X, Y, Z \in \text{Mod}\}$ *is a set of attack relation.*

4.4 Proof Theory

The formalism we have introduced allows us to temporalise rules, thus we have to admit the possibility that rules are not only given but can be proved to holf for certain span of time. Accordingly we have to give conditions that allow us to derive rules instead of literals. A conclusion of a theory D is a tagged temporal literal or rule having one of the following forms:

$+\Delta\gamma{:}V$ meaning that $\gamma{:}V$ is definitely provable in D.

$-\Delta\gamma{:}V$ meaning that $\gamma{:}V$ is not definitely provable in D.
$+\partial\gamma{:}V$ meaning that $\gamma{:}V$ is defeasible provable in D.
$-\partial\gamma{:}V$ meaning that $\gamma{:}V$ is not defeasible provable in D.

Provability is based on the concept of a derivation (or proof) in D. A derivation is a finite sequence $P = (P(1),..,P(n))$ of tagged literals. Each tagged temporal literal or rule satisfies some proof conditions, which correspond to inference rules for the four kinds of conclusions we have mentioned above. The conditions for applicability of rules are formalised below:

If Convert(Y,X) and $r{:}\overline{T_r}$ is Δ-*applicable* in the proof condition for $\pm\Delta_X$ then
(1) $+\Delta r{:}\overline{T_r} \in P(1..i)$, and either
(2) $r{:}\overline{T_r} \in R^X$,
  (2.1) $\forall\alpha{:}\overline{T_\alpha} \in A(r{:}\overline{T_r})$,
    (2.1.1) $+\Delta_{\text{KNOW}}\alpha : \overline{T_\alpha} \in P(1..i)$, or $+\Delta_{\text{KNOW}}\alpha : \overline{T_\alpha}{:}\widehat{T_r} \in P(1..i)$, or
    (2.1.2) $+\Delta_{\text{ACT}}\alpha : \overline{T_\alpha} \in P(1..i)$, or $+\Delta_{\text{ACT}}\alpha : \overline{T_\alpha}{:}\widehat{T_r} \in P(1..i)$,
  (2.2) $\forall\alpha{:}\widehat{T_\alpha} \in A(r{:}\overline{T_r})$,
    (2.2.1) $+\Delta_{\text{KNOW}}\alpha : \widehat{T_\alpha} \in P(1..i)$, or $+\Delta_{\text{KNOW}}\alpha : \widehat{T_\alpha}{:}\widehat{T_r} \in P(1..i)$, or
    (2.2.2) $+\Delta_{\text{ACT}}\alpha : \widehat{T_\alpha} \in P(1..i)$, or $+\Delta_{\text{ACT}}\alpha : \widehat{T_\alpha}{:}\widehat{T_r} \in P(1..i)$, and
  (2.3) $\forall Z\alpha{:}\overline{T_\alpha} \in A(r{:}\overline{T_r}), +\Delta_Z\alpha : \overline{T_\alpha} \in P(1..i)$, or $+\Delta_Z\alpha : \overline{T_\alpha}{:}\widehat{T_r} \in P(1..i)$, and
  (2.4) $\forall Z\alpha{:}\widehat{T_\alpha} \in A(r{:}\overline{T_r}), +\Delta_Z\alpha : \widehat{T_\alpha} \in P(1..i)$, or $+\Delta_Z\alpha : \widehat{T_\alpha}{:}\widehat{T_r} \in P(1..i)$, or
(3) $r{:}\overline{T_r} \in R^Y$,
  (3.1) $\forall\alpha{:}\overline{T_\alpha} \in A(r{:}\overline{T_r})$, $+\Delta_X\alpha : \overline{T_\alpha} \in P(1..i)$, or $+\Delta_X\alpha : \overline{T_\alpha}{:}\widehat{T_r} \in P(1..i)$, and
  (3.2) $\forall\alpha{:}\widehat{T_\alpha} \in A(r{:}\overline{T_r}), +\Delta_X\alpha : \widehat{T_\alpha} \in P(1..i)$, or $+\Delta_X\alpha : \widehat{T_\alpha}{:}\widehat{T_r} \in P(1..i)$.

The conditions for a rule r to be ∂-*applicable* are the same as those for Δ-*applicable*, but where we replace Δ with ∂.

If Convert(Y,X) and $r{:}\overline{T_r}$ is Δ-*discarded* in the proof condition for $\pm\Delta_X$ then
(1) $-\Delta r{:}\overline{T_r} \in P(1..i)$, or either
(2) $r{:}\overline{T_r} \in R^X$,
  (2.1) $\exists\alpha{:}\overline{T_\alpha} \in A(r{:}\overline{T_r})$,
    (2.1.1) $-\Delta_{\text{KNOW}}\alpha : \overline{T_\alpha} \in P(1..i)$, and $-\Delta_{\text{KNOW}}\alpha : \overline{T_\alpha}{:}\widehat{T_r} \in P(1..i)$, and
    (2.1.2) $-\Delta_{\text{ACT}}\alpha : \overline{T_\alpha} \in P(1..i)$, and $-\Delta_{\text{ACT}}\alpha : \overline{T_\alpha}{:}\widehat{T_r} \in P(1..i)$, or
  (2.2) $\exists\alpha{:}\widehat{T_\alpha} \in A(r{:}\overline{T_r})$,
    (2.2.1) $-\Delta_{\text{KNOW}}\alpha : \widehat{T_\alpha} \in P(1..i)$, and $-\Delta_{\text{KNOW}}\alpha : \widehat{T_\alpha}{:}\widehat{T_r} \in P(1..i)$, and
    (2.2.2) $-\Delta_{\text{ACT}}\alpha : \widehat{T_\alpha} \in P(1..i)$, and $-\Delta_{\text{ACT}}\alpha : \widehat{T_\alpha}{:}\widehat{T_r} \in P(1..i)$, or
  (2.3) $\exists Z\alpha{:}\overline{T_\alpha} \in A(r{:}\overline{T_r}), -\Delta_Z\alpha : \overline{T_\alpha} \in P(1..i)$, and $-\Delta_Z\alpha : \overline{T_\alpha}{:}\widehat{T_r} \in P(1..i)$, or
  (2.4) $\exists Z\alpha{:}\widehat{T_\alpha} \in A(r{:}\overline{T_r}), -\Delta_Z\alpha : \widehat{T_\alpha} \in P(1..i)$, and $-\Delta_Z\alpha : \widehat{T_\alpha}{:}\widehat{T_r} \in P(1..i)$, or
(3) $r{:}\overline{T_r} \in R^Y$,
  (3.1) $\exists\alpha{:}\overline{T_\alpha} \in A(r{:}\overline{T_r})$, $-\Delta_X\alpha : \overline{T_\alpha} \in P(1..i)$, and $-\Delta_X\alpha : \overline{T_\alpha}{:}\widehat{T_r} \in P(1..i)$, or
  (3.2) $\exists\alpha{:}\widehat{T_\alpha} \in A(r{:}\overline{T_r}), -\Delta_X\alpha : \widehat{T_\alpha} \in P(1..i)$, and $-\Delta_X\alpha : \widehat{T_\alpha}{:}\widehat{T_r} \in P(1..i)$.

The conditions for a rule $r{:}\overline{T_r}$ to be ∂-*discarded* are the same as those for Δ-*discarded*, but where we replace Δ with ∂.

We are now ready to define the proof theory that is, the inference conditions to derive tagged conclusions from a given theory D. We begin with the proof conditions to determine whether a rule is a definite conclusion of a theory D. A temporalised rule $\gamma{:}\overline{T}$ is definitely provable ($+\Delta$) if (1) there exists a rule $\gamma{:}\overline{T_\gamma}$ in the set of rule such that $\overline{T} \sqsubseteq \overline{T_\gamma}$, or (2) blalblabla Formally:

If $P(i+1) = +\Delta\,\gamma{:}\overline{T}$ then
(1) $\exists\overline{T_\gamma}, \sqsubseteq \overline{T_\gamma}, \gamma{:}\overline{T_\gamma} \in R, \overline{T}$, or
(2) $\exists\overline{T_{\gamma 1}}, \exists\overline{T_{\gamma 2}}$, meets($\overline{T_{\gamma 1}}, \overline{T_{\gamma 2}}$), start($\overline{T_{\gamma 1}}$) = start($\overline{T}$), end($\overline{T_{\gamma 2}}$) = end($\overline{T}$),
$\gamma{:}\overline{T_{\gamma 1}} \in R$ and $\gamma{:}\overline{T_{\gamma 2}} \in R$.

A definite rule is not provable at interval $\overline{T}$ if there is not such rule in the set rules defined in a larger interval.

If $P(i+1) = -\Delta\,\gamma{:}\overline{T}$ then
(1) $\forall\overline{T_\gamma}, \sqsubseteq \overline{T_\gamma}, \gamma{:}\overline{T_\gamma} \notin R, \overline{T}$, and
(2) $\forall\overline{T_{\gamma 1}}, \forall\overline{T_{\gamma 2}}$, meets($\overline{T_{\gamma 1}}, \overline{T_{\gamma 2}}$), start($\overline{T_{\gamma 1}}$) = start($\overline{T}$), end($\overline{T_{\gamma 2}}$) = end($\overline{T}$),
$\gamma{:}\overline{T_{\gamma 1}} \notin R$ or $\gamma{:}\overline{T_{\gamma 2}} \notin R$.

A temporalised rule $\gamma{:}\widehat{T}$ is definitely provable ($+\Delta$) if there exists a rule $\gamma{:}\overline{T_\gamma}$ in the set of rule such that $\overline{T} \sqsubseteq \overline{T_\gamma}$. Formally:

If $P(i+1) = +\Delta\,\gamma{:}\widehat{T}$ then $\exists\overline{T_\gamma}$,over$(\widehat{T}, \overline{T_\gamma}), \gamma{:}\overline{T_\gamma} \in R$.

If $P(i+1) = -\Delta\,\gamma{:}\widehat{T}$ then $\forall\overline{T_\gamma}$,over$(\widehat{T}, \overline{T_\gamma}), \gamma{:}\overline{T_\gamma} \notin R$.

We can now move to definite literals.

If $P(i+1) = +\Delta_X\,\gamma{:}\overline{V}$ and Convert(Y, X) then
(1) $\exists\overline{T_\gamma}, \overline{V} \sqsubseteq \overline{T_\gamma}, X\gamma{:}\overline{T_\gamma} \in F$, or
(2) if $X =$ KNOW then $\exists\overline{T_\gamma}, \overline{V} \sqsubseteq \overline{T_\gamma}, \gamma{:}\overline{T_\gamma} \in F$, or
(3) if $X =$ INT then $\exists\overline{V_\gamma}, \overline{V} \sqsubseteq \overline{V_\gamma}, +\Delta_{\mathrm{ACT}}\gamma{:}\overline{V_\gamma}$, or
(4) $\exists r{:}\overline{T_r} \in R_s[\gamma : \overline{T_\gamma}], \overline{V} \sqsubseteq \overline{T_\gamma}{:}\overline{T_r}$, $r{:}\overline{T_r}$ is Δ-*applicable*, or
(5) $\exists\overline{V_{\gamma 1}}, \exists\overline{V_{\gamma 2}}$, meets($\overline{V_{\gamma 1}}, \overline{V_{\gamma 2}}$), start($\overline{V_{\gamma 1}}$) = start($\overline{V}$), end($\overline{V_{\gamma 2}}$) = end($\widehat{V}$)
$+\Delta_X\,\gamma{:}\overline{V_{\gamma 1}} \in P(1..i)$ and $+\Delta_X\,\gamma{:}\overline{V_{\gamma 2}} \in P(1..i)$.

To prove that a definite literal is not provable we have to show that all attempts to give a definite proof of the literal fail.

If $P(i+1) = -\Delta_X\,\gamma{:}\overline{V}$ and Convert(Y, X) then
(1) $\forall\overline{T_\gamma}, \overline{V} \sqsubseteq \overline{T_\gamma}, X\gamma{:}\overline{T_\gamma} \notin F$, and
(2) if $X =$ KNOW then $\forall\overline{T_\gamma}, \overline{V} \sqsubseteq \overline{T_\gamma}, \gamma{:}\overline{T_\gamma} \notin F$, and
(3) if $X =$ INT then $\forall\overline{V_\gamma}, \overline{V} \sqsubseteq \overline{V_\gamma}, -\Delta_{\mathrm{ACT}}\gamma{:}\overline{V_\gamma}$, and
(4) $\forall r{:}\overline{T_r} \in R_s[\gamma : \overline{T_\gamma}], \overline{V} \sqsubseteq \overline{T_\gamma}{:}\overline{T_r}$, $r{:}\overline{T_r}$ is Δ-*discarded*, and
(5) $\forall\overline{V_{\gamma 1}}, \forall\overline{V_{\gamma 2}}$, meets($\overline{V_{\gamma 1}}, \overline{V_{\gamma 2}}$), start($\overline{V_{\gamma 1}}$) = start($\overline{V}$), end($\overline{V_{\gamma 2}}$) = end($\widehat{V}$)
$-\Delta_X\,\gamma{:}\overline{V_{\gamma 1}} \in P(1..i)$ or $-\Delta_X\,\gamma{:}\overline{V_{\gamma 2}} \in P(1..i)$.

The conditions for a temporal literal $\gamma{:}\widehat{T}$ to be definitely provable ($+\Delta$) are formally expressed below.

If $P(i+1) = +\Delta_X \gamma{:}\widehat{V}$ and $\mathrm{Convert}(Y,X)$ then $\exists \widehat{V_\gamma}, over(\widehat{V}, \overline{V_\gamma}), +\Delta_X \gamma{:}\overline{V_\gamma}$.

If $P(i+1) = -\Delta_X \gamma{:}\widehat{V}$ and $\mathrm{Convert}(Y,X)$ then $\forall \widehat{V_\gamma}, over(\widehat{V}, \overline{V_\gamma}), -\Delta_X \gamma{:}\overline{V_\gamma}$.

If $P(i+1) = +\Delta_X \gamma{:}\overline{T}{:}\widehat{T_r}$ and $\mathrm{Convert}(Y,X)$ then
$\exists \overline{T_{\gamma 1}}, \exists \overline{T_{r1}}, \overline{T} \sqsubseteq \overline{T_{\gamma 1}}$, over$(\widehat{T_r}, \overline{T_{r1}}), +\Delta_X \gamma{:}\overline{T_{\gamma 1}}{:}\overline{T_{r1}} \in$P(1..i).

If $P(i+1) = -\Delta_X \gamma{:}\overline{T}{:}\widehat{T_r}$ and $\mathrm{Convert}(Y,X)$ then
$\forall \overline{T_{\gamma 1}}, \forall \overline{T_{r1}}, \overline{T} \sqsubseteq \overline{T_{\gamma 1}}$, over$(\widehat{T_r}, \overline{T_{r1}}), -\Delta_X \gamma{:}\overline{T_{\gamma 1}}{:}\overline{T_{r1}} \in$P(1..i).

We now turn our attention to defeasible derivations, that is, derivations giving a temporal assertion $\gamma{:}V$ as a defeasible conclusion of a theory D. We begin with the proof conditions to determine whether a rule is a defeasible conclusion.

If $P(i+1) = +\partial r{:}\overline{T}$ then $+\Delta r{:}\overline{T} \in P(1..i)$

If $P(i+1) = +\partial r{:}\widehat{T}$ then $+\Delta r{:}\widehat{T} \in P(1..i)$.

Defeasible provability ($+\partial$) for temporal literals consists of three phases. In the first phase, we put forward a supported reason for the temporal assertion that we want to prove. Then in the second phase, we consider all possible attacks against the desired conclusion. Finally in the last phase, we have to counter-attack the attacks considered in the second phase.

If $P(i+1) = +\partial_X \gamma{:}\overline{V}$ and $\mathrm{Convert}(Y,X)$ and $\mathrm{Attack}(W,Z,X)$ then
(1) $+\Delta_X \gamma{:}\overline{V} \in P(1..i)$, or
(2) $-\Delta_X {\sim}\gamma{:}\widehat{V} \in P(1..i)$, and
  (2.1) if $X = \mathrm{INT}$ then $\exists \overline{V_\gamma}, \overline{V} \sqsubseteq \overline{V_\gamma}, +\partial_{\mathrm{ACT}} \gamma{:}\overline{V_\gamma}$, or
  (2.2) $\exists r{:}\overline{T_r} \in R_{sd}^{X \cup Y}[\gamma{:}\overline{T_\gamma}]$, $\overline{V} \sqsubseteq \overline{T_\gamma}{:}\overline{T_r}$, $r{:}\overline{T_r}$ is ∂-applicable,
  (2.3) $\forall s{:}\overline{T_s} \in R^{W \cup Z \cup X \cup Y}[{\sim}\gamma{:}\overline{T_{\sim\gamma}}]$, si$(\overline{T_{\sim\gamma}}{:}\overline{T_s}, \overline{T_\gamma}{:}\overline{T_r})$, sbe$(\overline{T_{\sim\gamma}}{:}\overline{T_s}, \overline{V})$,
    (2.3.1) $s{:}\overline{T_s}$ is ∂-discarded, or
    (2.3.2) $\exists w{:}\overline{T_w} \in R^K[\gamma{:}\overline{T_{w\gamma}}]$, $\overline{V} \sqsubseteq \overline{T_{w\gamma}}{:}\overline{T_w}$,
      (2.3.2.1) $w{:}\overline{T_w}$ is $+\partial$-applicable, and either
      (2.3.2.2) $s{:}\overline{T_s} \in R^{X \cup Y}$,
        (2.3.2.2.1) $w{:}\overline{T_w} \in R^{W \cup Z}$, or
        (2.3.2.2.2) $w{:}\overline{T_w} \in R^{X \cup Y}$, $w{:}\overline{T_w} \succ s{:}\overline{T_s}$, or
      (2.3.2.3) $s{:}\overline{T_s} \in R^Z$,
        (2.3.2.3.1) $w{:}\overline{T_w} \in R^W$, or
        (2.3.2.3.2) $w{:}\overline{T_w} \in R^Z$, $w{:}\overline{T_w} \succ s{:}\overline{T_s}$, or
      (2.3.2.4) $s{:}\overline{T_s} \in R^W$, $w{:}\overline{T_w} \in R^W$, $w{:}\overline{T_w} \succ s{:}\overline{T_s}$, or
(3) $\exists \overline{V_{\gamma 1}}, \exists \overline{V_{\gamma 2}}$, meets$(\overline{V_{\gamma 1}}, \overline{V_{\gamma 2}})$, start$(\overline{V_{\gamma 1}})$ = start$(\overline{V})$, end$(\overline{V_{\gamma 2}})$ = end$(\widehat{V})$
$+\partial_X \gamma{:}\overline{V_{\gamma 1}} \in P(1..i)$ and $+\partial_X \gamma{:}\overline{V_{\gamma 2}} \in P(1..i)$.

If $P(i+1) = -\partial_X \gamma{:}\overline{V}$ and $\mathrm{Convert}(Y,X)$ and $\mathrm{Attack}(W,Z,X)$ then
(1) $-\Delta_X \gamma{:}\overline{V} \in P(1..i)$, and
(2) $+\Delta_X \sim\gamma{:}\widehat{V} \in P(1..i)$, or
(2.1) if $X = \mathrm{INT}$ then $\forall \overline{V_\gamma}, \overline{V} \sqsubseteq \overline{V_\gamma}$, $-\partial_{\mathrm{ACT}}\gamma{:}\overline{V_\gamma}$, and
(2.2) $\forall r{:}\overline{T_r} \in R_{sd}^{X\cup Y}[\gamma{:}\overline{T_\gamma}]$, $\overline{V} \sqsubseteq \overline{T_\gamma}{:}\overline{T_r}$, $r{:}\overline{T_r}$ is ∂-applicable,
(2.3) $\exists s{:}\overline{T_s} \in R^{W\cup Z\cup X\cup Y}[\sim\gamma{:}\overline{T_{\sim\gamma}}]$, $\mathrm{si}(\overline{T_{\sim\gamma}}{:}\overline{T_s}, \overline{T_\gamma}{:}\overline{T_r})$, $\mathrm{sbe}(\overline{T_{\sim\gamma}}{:}\overline{T_s}, \overline{V})$,
(2.3.1) $s{:}\overline{T_s}$ is ∂-applicable, and
(2.3.2) $\forall w{:}\overline{T_w} \in R[\gamma{:}\overline{T_{w\gamma}}]$, $\overline{T} \sqsubseteq \overline{T_{w\gamma}}{:}\overline{T_w}$, either
(2.3.2.1) $w{:}\overline{T_w}$ is ∂-discarded, or
(2.3.2.2) $s{:}\overline{T_s} \in R^{X\cup Y}$,
(2.3.2.2.1) $w{:}\overline{T_w} \notin R^{W\cup Z}$, and
(2.3.2.2.2) $w{:}\overline{T_w} \in R^{X\cup Y}$, $w{:}\overline{T_w} \not> s{:}\overline{T_s}$, and
(2.3.2.3) $s{:}\overline{T_s} \in R^Z$
(2.3.2.3.1) $w{:}\overline{T_w} \notin R^W$, and
(2.3.2.3.2) $w{:}\overline{T_w} \in R^Z$, $w{:}\overline{T_w} \not> s{:}\overline{T_s}$, and
(2.3.2.4) $s{:}\overline{T_s} \in R^W$, $w{:}\overline{T_w} \in R^W$, $w{:}\overline{T_w} \not> s{:}\overline{T_s}$, and
(3) $\forall \overline{V_{\gamma 1}}, \forall \overline{V_{\gamma 2}}$, $\mathrm{meets}(\overline{V_{\gamma 1}}, \overline{V_{\gamma 2}})$, $\mathrm{start}(\overline{V_{\gamma 1}}) = \mathrm{start}(\overline{V})$, $\mathrm{end}(\overline{V_{\gamma 2}}) = \mathrm{end}(\widehat{V})$
$-\partial_X \gamma{:}\overline{V_{\gamma 1}} \in P(1..i)$ and $-\partial_X \gamma{:}\overline{V_{\gamma 2}} \in P(1..i)$

Let us illustrate the proof condition of the defeasible provability of X$\gamma{:}\overline{V}$. We have two cases: 1) We show that X$\gamma{:}\overline{V}$ is already definitely provable; or 2) we need to argue using the defeasible part of D. In this second case, to prove X$\gamma{:}\overline{V}$ defeasibly we must show that X$\sim\gamma{:}\widehat{V}$ is not definitely provable (2). We require then there must be a strict or defeasible rule $r{:}\overline{T_r} \in R^{X\cup Y}$ which can be applied and with head $\gamma{:}\overline{T_\gamma}$ such that $\overline{V} \sqsubseteq \overline{T_\gamma}{:}\overline{T_r}$ (2.1). But now we need to consider possible attacks, i.e., reasoning chains in support of $\sim\gamma{:}\overline{V}$, that is, any rule $s{:}\overline{T_s} \in R^{W\cup Z\cup X\cup Y}$ which has head $\sim\gamma{:}\overline{T_{\sim\gamma}}$ such that $\mathrm{si}(\overline{T_{\sim\gamma}}{:}\overline{T_s}, \overline{T_\gamma}{:}\overline{T_r})$, and $\mathrm{sbe}(\overline{T_{\sim\gamma}}{:}\overline{T_s}, \overline{V})$. Note that here we consider defeaters, too, whereas they could not be used to support the conclusion X$\gamma{:}\overline{V}$; this is in line with the motivation of defeaters given earlier. These attacking rules $s{:}T_s$ have to be discarded (2.3.1), or must be counterattacked by a stronger rule $w{:}T_w$ which has a head $\gamma{:}T_{w\gamma}$ such that $\overline{V}$ is contained in $\overline{T_{w\gamma}}{:}\overline{T_w}$ (2.3.2). blabla

The defeasible proof for a temporalised literal to hold in some instants of a chain of viewpoints $\widehat{V}$

If $P(i+1) = +\partial_X \gamma{:}\widehat{V}$ and $\mathrm{Convert}(Y,X)$ and $\mathrm{Attack}(W,Z,X)$ then
$\exists \overline{V_\gamma}$,over$(\widehat{V}, \overline{V_\gamma}), +\partial_X \gamma{:}\overline{V_\gamma} \in$P(1..i).

If $P(i+1) = -\partial_X \gamma{:}\widehat{V}$ and $\mathrm{Convert}(Y,X)$ and $\mathrm{Attack}(W,Z,X)$ then
$\forall \overline{V_\gamma}$,over$(\widehat{V}, \overline{V_\gamma}), -\partial_X \gamma{:}\overline{V_\gamma} \in$P(1..i).

5 Conclusions

We hope you like the Dagstuhl Seminar Proceedings :-)
pricai06

References

1. Governatori, G., Padmanabhan, V., Rotolo, A.: Rule-based agents in temporalised defeasible logic. In Yang, Q., Webb, G., eds.: Ninth Pacific Rim International Conference on Artificial Intelligence. Number 4099 in LNAI, Berlin, Springer (2006) 31–40
2. Dastani, M., van der Torre, L.W.N.: Programming boid-plan agents: Deliberating about conflicts among defeasible mental attitudes and plans. In: AAMAS. (2004) 706–713
3. Broersen, J., Dastani, M., Hulstijn, J., van der Torre, L.: Goal generation in the BOID architecture. Cognitive Science Quarterly **2** (2002) 428–447
4. Dastani, M., de Boer, F., Dignum, F., Meyer, J.J.: Programming agent deliberation: an approach illustrated using the 3apl language. In: AAMAS '03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems, New York, NY, USA, ACM Press (2003) 97–104
5. Dastani, M., van Riemsdijk, B., Dignum, F., Meyer, J.: A programming language for cognitive agents: Goal-directed 3apl. In: PROMAS. (2003) 111–130
6. van Riemsdijk, M.B., Dastani, M., Meyer, J.J.C.: Semantics of declarative goals in agent programming. In: AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems, New York, NY, USA, ACM Press (2005) 133–140
7. Governatori, G., Rotolo, A.: Defeasible logic: Agency, intention and obligation. In Lomuscio, A., Nute, D., eds.: Deontic Logic in Computer Science. Number 3065 in LNAI, Berlin, Springer-Verlag (2004) 114–128
8. Dastani, M., Governatori, G., Rotolo, A., van der Torre, L.: Programming cognitive agents in defeasible logic. In Sutcliffe, G., Voronkov, A., eds.: Proc. LPAR 2005. Volume 3835 of LNAI., Springer (2005) 621–636
9. Dastani, M., Governatori, G., Rotolo, A., van der Torre, L.: Preferences of agents in defeasible logic. In Zhang, S., Jarvis, R., eds.: Proc. Australian AI05. Volume 3809 of LNAI., Springer (2005) 695–704
10. Governatori, G., Rotolo, A., Sartor, G.: Temporalised normative positions in defeasible logic. In Gardner, A., ed.: 10th International Conference on Artificial Intelligence and Law (ICAIL05), ACM Press (2005) 25–34
11. Governatori, G., Palmirani, M., Riveret, R., Rotolo, A., Sartor, G.: Norm modifications in defeasible logic. In Moens, M.F., Spyns, P., eds.: Legal Knowledge and Information Systems. Number 134 in Frontieres in Artificial Intelligence and Applications. IOS Press, Amsterdam (2005) 13–22
12. Bratman, M.E.: Intentions, Plans and Practical Reason. Harvard University Press, Cambridge, MA (1987)
13. Palmirani, M., Brighi, R.: Time model for managing the dynamic of normative system. In: Proceedings of EGOV 2006, Berlin, Springer (2006) 207–218
14. Nute, D.: Defeasible reasoning. In: Proceedings of 20th Hawaii International Conference on System Science, IEEE press (1987)
15. Nute, D.: Defeasible logic. In Gabbay, D., Hogger, C., Robinson, J., eds.: Handbook of Logic in Artificial Intelligence and Logic Programming. Volume 3. Oxford University Press (1993) 353–395
16. Antoniou, G., Billington, D., Governatori, G., Maher, M.J.: Representation results for defeasible logic. ACM Transactions on Computational Logic **2** (2001) 255–287
17. Governatori, G., Maher, M.J., Billington, D., Antoniou, G.: Argumentation semantics for defeasible logics. Journal of Logic and Computation **14** (2004) 675–702
18. Maher, M.J.: Propositional defeasible logic has linear complexity. Theory and Practice of Logic Programming **1** (2001) 691–711

19. Bassiliades, N., Antoniou, G., Vlahavas, I.: DR-DEVICE: A defeasible logic system for the Semantic Web. In Ohlbach, H.J., Schaffert, S., eds.: 2nd Workshop on Principles and Practice of Semantic Web Reasoning. Number 3208 in LNCS, Springer (2004) 134–148